WHAT IS CLAIMED IS:

1. A computer-based method for modeling complex formation between a query ligand and a target macromolecule, the method comprising:

a) providing a structural model of a query ligand and a structural model of a target macromolecule;

b) identifying a substructure of the query ligand;

c) identifying comparison ligands in a set of 3-D structural models that each share an identical substructure with the query ligand, wherein each 3-D structural model comprises a comparison ligand and a comparison macromolecule, and wherein the comparison macromolecule has structural features homologous to the target macromolecule;

d) mapping spatial relationships between the substructure atoms of the query ligand and the comparison ligand such that corresponding atoms are identified;

e) assigning atomic coordinates to the corresponding atoms of the query ligand;

f) generating one or more output models, each model comprising a 3-D structural model of the query ligand substructure and the target macromolecule, wherein the 3-D model of the query ligand substructure comprises the atomic coordinates of the query ligand from step (e).

2. The method of claim 1, wherein the query ligand is less than 1000 Daltons MW.

3. The method of claim 1, wherein the query ligand is an inhibitor of the target macromolecule.

4. The method of claim 1, wherein the query ligand is an inhibitor of the comparison macromolecule.

5. The method of claim 1, wherein the output models comprise models in which non-substructure atoms of the query ligand are represented .

6. The method of claim 1, wherein a plurality of query ligands are provided.

7. The method of claim 1, wherein the substructure comprises 2-D structural information.

8. The method of claim 7, wherein the substructure comprises a framework.

9. The method of claim 8, wherein the framework comprises cyclic atoms of the query ligand, acyclic atoms that connect the cyclic portions, and sp^2-hybridized oxygen atoms connected to the cyclic and acyclic atoms.

10. The method of claim 7, wherein the substructure comprises a substructure in which at least 5, 7, or 10 atoms are identical in the comparison ligand(s).

11. The method of claim 1, wherein the substructure comprises 3-D structural information.

12. The method of claim 1, wherein the substructure comprises a pharmacophore.

13. The method of claim 12, wherein the identifying the pharmacophore comprises identifying comparison ligand atoms which form hydrogen-bonds with a macromolecule of interest.

14. The method of claim 13, wherein the macromolecule of interest is the comparison macromolecule.

15. The method of claim 1, wherein the target macromolecule and the comparison macromolecule are identical.

16. The method of claim 1, further comprising refining the output models.

17. The method of claim 1, wherein the target macromolecule is a polypeptide or a nucleic acid.

18. The method of claim 16, wherein the refining comprises performing rigid body minimization or minimization with flexible ligand sidechains.

19. The method of claim 17, wherein each output model comprises the 3-D spatial positions of amino acid backbone C and N atoms of the target macromolecule.

20. The method of claim 19, wherein each output model comprises the 3-D spatial positions of amino acid backbone Cα atoms of the target macromolecule.

21. The method of claim 17, wherein each output model comprises the 3-D spatial positions of amino acid sidechain C, N, S, and O atoms of the target macromolecule.

22. The method of claim 17, wherein each output model comprises the 3-D spatial positions of H atoms of the target macromolecule.

23. The method of claim 22, wherein each output model comprises the 3-D spatial positions of polar H atoms.

24. The method of claim 6, further comprising evaluating each output model of the plurality.

25. The method of claim 24, wherein the evaluating comprises determining one or more of lipophilic interactions, hydrogen bonding, repulsion, and intramolecular strain energy between the substructure and target macromolecule.

26. The method of claim 25, further comprising assigning a score to each output model.

27. The method of claim 26, further comprising obtaining physical samples comprising a subset of the query ligands, wherein the ligands of the subset are assigned a preselected score.

28. The method of claim 27, further comprising evaluating the binding of the ligands of the subset to the target macromolecule.

29. The method of claim 1, wherein the set of 3-D structural models is contained in a database.

30. An apparatus comprising:

a) a memory that stores executable instructions for modeling complex formation between a query ligand and a target macromolecule, and

b) a processor that executes the instructions to:

i) provide a structural model of a query ligand and a target macromolecule;

ii) identify a substructure of the query ligand;

iii) identify comparison ligands in a set of 3-D structural models that each share an identical substructure with the query ligand, wherein each 3-D structural model comprises a comparison ligand and a comparison macromolecule, and wherein the comparison macromolecule has structural features homologous to the target macromolecule;

iv) map spatial relationships between the substructure atoms of the query ligand and the comparison ligand such that corresponding atoms are identified;

v) assign atomic coordinates to the corresponding atoms of the query ligand;

vi) generate one or more output models, each model comprising a 3-D structural model of the query ligand substructure and the target macromolecule, wherein the 3-D model of the query ligand substructure comprises the atomic coordinates of the query ligand from step (v).

123

124 31. An article comprising machine-readable media that stores executable instructions
125 for modeling complex formation between a query ligand and a target macromolecule, the
126 instructions causing a machine to:
127 a) provide a structural model of a query ligand and a target macromolecule;
128 b) identify a substructure of the query ligand;
129 c) identify comparison ligands in a set of 3-D structural models that each
130 share an identical substructure with the query ligand, wherein each 3-D structural model
131 comprises a comparison ligand and a comparison macromolecule, and wherein the
132 comparison macromolecule has structural features homologous to the target macromolecule;
133 d) map spatial relationships between the substructure atoms of the query
134 ligand and the comparison ligand such that corresponding atoms are identified;
135 e) assign atomic coordinates to the corresponding atoms of the query ligand;
136 f) generate one or more output models, each model comprising a 3-D
137 structural model of the query ligand substructure and the target macromolecule, wherein the
138 3-D model of the query ligand substructure comprises the atomic coordinates of the query
139 ligand from step (e).
140

141 32. A database of ligand-protein structure models, the database comprising a
142 plurality of records, each record comprising information representing 3-D spatial positions of
143 atoms in a protein and atoms in a ligand that physically interacts with the protein, wherein
144 the database includes at least two classes of records:
145 a) a first class for which the 3-D spatial positions of atoms of each model are
146 determined by a physical observation; and
147 b) a second class for which the 3-D spatial positions of atoms of each model
148 of the set are inferred by the following steps:
149 i) identifying models from the first class that comprise a ligand having
150 a substructure identical to a query ligand, and having a protein that comprises
151 structural features homologous to a target protein;

ii) mapping spatial relationships between the substructure atoms of the query ligand and the comparison ligand such that corresponding atoms are identified;

iii) assigning atomic coordinates to the corresponding atoms of the query ligand;

iv) generating one or more output models, each model comprising a 3-D structural model of the query ligand substructure and the target macromolecule, wherein the 3-D model of the query ligand substructure comprises the atomic coordinates of the query ligand from step (iv).

33. The database of claim 32, further comprising a third class of records, for which the 3-D spatial positions of atoms of each model of the set are inferred by the following steps:

vi) providing the output models of the second class;

vii) modifying the substructure to comprise one or more additional atoms of the query ligand.

34. A computer-based method for modeling complex formation between a test ligand and a target macromolecule, the method comprising:

a) providing a 3-D structural model of a ligand and a target macromolecule;

b) identifying a substructure of the compound;

c) identifying test ligands in a set of structural models that each share an identical substructure with the compound;

d) mapping spatial relationships between the substructure atoms of the ligand and the test ligand such that corresponding atoms of the test ligand are identified;

e) assigning atomic coordinates to the corresponding atoms of the test ligand;

f) generating one or more output models, each model comprising a 3-D structural model of the test ligand and the target macromolecule, wherein the 3-D model of the test ligand comprises the atomic coordinates of the test ligand from step (e), thereby modeling complex formation between a test ligand and a target macromolecule.